

Compaction of Duplex Nucleic Acids upon Native Electrospray Mass Spectrometry

Massimiliano Porrini,¹ Frédéric Rosu,² Clémence Rabin,¹ Leonardo Darré,^{3,4} Hansel Gómez,^{3,4} Modesto Orozco,^{3,4,5} and Valérie Gabelica*,¹

¹ Univ. Bordeaux, INSERM, CNRS, Laboratoire Acides Nucléiques Régulations Naturelle et Artificielle (ARNA, U1212, UMR5320), IECB, 2 rue Robert Escarpit, 33607 Pessac, France.

² CNRS, INSERM, Univ. Bordeaux, Institut Européen de Chimie et Biologie (IECB, UMS3033, US001), 2 rue Robert Escarpit, 33607 Pessac, France.

³ Institute for Research in Biomedicine (IRB) Barcelona, the Barcelona Institute of Science and Technology, Barcelona, Spain.

⁴ Joint BSC-IRB Research Program in Computational Biology, IRB Barcelona, Barcelona, Spain.

⁵ Department of Biochemistry and Biomedicine, University of Barcelona, Barcelona, Spain.

Corresponding Author: v.gabelica@iecb.u-bordeaux.fr

ABSTRACT:

Native mass spectrometry coupled to ion mobility spectrometry is a promising tool for structural biology. Intact complexes can be transferred to the mass spectrometer and, if native conformations survive, collision cross sections give precious information on the structure of each species in solution. Based on several successful reports for proteins and their complexes, the conformation survival becomes more and more taken for granted. Here we report on the fate of nucleic acids conformation in the gas phase. Disturbingly, we found that DNA and RNA duplexes, at the electrospray charge states naturally obtained from native solution conditions (≥ 100 mM aqueous NH_4OAc), are significantly more compact in the gas phase compared to the canonical solution structures. The compaction is observed for short (12-bp) and long (36-bp) duplexes, and for DNA and RNA alike. Molecular modeling (density functional calculations on small helices, semi-empirical calculations on up to 12-bp, and molecular dynamics on up to 36-bp duplexes) demonstrates that the compaction is due to phosphate group self-solvation prevailing over Coulomb-driven expansion. Molecular dynamics simulations starting from solution structures do not reproduce the experimental compaction. To be experimentally relevant, molecular dynamics sampling should reflect the progressive structural rearrangements occurring during desolvation. For nucleic acid duplexes, the compaction observed for low charge states results from novel phosphate-phosphate hydrogen bonds formed across both grooves at the very late stages of electrospray.

INTRODUCTION

Besides genetic information storage, nucleic acids perform pivotal regulatory functions in living organisms.¹ Conformational changes are key to these regulation mechanisms, whether at the DNA level (e.g., particular structures in gene promoters)² or at the RNA level (e.g., riboswitches).³ Electrospray ionization mass spectrometry (ESI-MS) in native conditions helps deciphering the complexation equilibria.⁴⁻⁷ Coupled to ion mobility spectrometry (IMS), native mass spectrometry makes it possible to reveal the ensuing conformational responses.⁸⁻¹⁰ But before native ESI-IMS-MS can be applied to study nucleic acids conformations from solution, it is essential to understand to what extent the different types of DNA/RNA secondary structures are preserved, or affected, by the transition from the solution to the gas phase.

Tandem mass spectrometry¹¹⁻¹⁴ experiments have shown that the kinetic stability of gas-phase DNA duplexes was correlated with the fraction of guanine-cytosine (GC) base pairs, and with the base pair ordering. This suggested that *Watson-Crick* (WC) hydrogen bonding and base stacking were at least partially maintained in the gas phase. Infrared multiphoton dissociation spectroscopy on duplexes and single strands suggested that GC bases are engaged in hydrogen bonds in the gas-phase duplexes, but no such evidence was found for AT base pairs.¹⁵ Early molecular dynamics (MD) validated this view: although (12-mer)₂⁶⁻ and (16-mer)₂⁸⁻ duplexes distort in the gas phase, most WC H-bonding and stacking interactions are preserved, particularly in GC-rich regions.¹⁶

The Bowers group later explored the gas-phase structure of DNA duplexes by IMS.^{17,18} By comparing the experimental collision cross section (CCS) of d(GC)_n duplexes with those obtained on duplexes relaxed by short (5-ns) MD, they showed that the gas-phase structures resemble an A-helix for the short duplexes (8—16-mer) and a B-helix for longer ones (>18-mer). It is worth noting that Bowers' IMS experiments were performed in non-native solution conditions (49:49:2 mixture of H₂O:MeOH:NH₄OH) and on high charge states (1 negative charge per 2 base pairs), raising questions on whether this represents what happens in native conditions. Here we show that, at the lower charge states

typically obtained from native solution conditions (100—150 mM NH₄OAc in 100% H₂O, pH=7), the conformations are significantly more compact than expected for B- or A-helices. We investigated the physical origin of this gas-phase compaction by molecular modeling.

METHODS

Electrospray ion mobility spectrometry. DNA and RNA duplexes were prepared by annealing their corresponding single-strands (purchased from Eurogentec, Seraing, Belgium, with RPl cartridge-Gold purification) in aqueous 100 mM NH₄OAc. When sprayed at 10 μM duplex, the major charge states are 4- for the 10-bp, 5- for the 12-bp, 7- for the 24-bp, 8- and 9- for the 36-bp duplexes. Higher charge states were generated by adding 0.2% to 0.75% sulfolane to the solution. ESI-IMS-MS experiments were recorded on an Agilent 6560 IMS-Q-TOF, with the drift tube operated in helium (supporting information Section S1). The arrival time distributions were fitted by Gaussian peaks and the CCS values of the center of each peak were determined by the stepped-field method. For visualization, we converted the arrival time distributions into CCS distributions (see supporting information).

Gas-phase simulations. The starting structures of the duplexes were built with the Nucleic Acid Builder (NAB) software,¹⁹ both for the A-form (DNA and RNA) and the B-form (DNA). Table 1 lists the main sequences and levels of theory used here. The numbers (33, 66, and 100) in duplex names reflect the GC content (in %) of each 12-bp unit. For 12-d₃₃, 12-d₆₆, 12-d₁₀₀ (B-helix) and 12-r₆₆ (A-helix), we first carried out MD simulations in water. Then all water molecules and counterions were removed at once, before each 1-μs gas-phase simulation.

The two possibilities to reduce the total charge to -5 (major charge state) are the localized charges (LC) and distributed charges (DC) models.¹⁶ LC and DC gave similar results upon unbiased MD of duplexes,¹⁶ so here only the LC model was tested extensively. With the LC model, protons are added on 17 out of the 22 phosphate groups. Among the 26334 possible protonation schemes, we selected a few low-energy ones based on single point molecular mechanics calculations. With

the DC model, the net charge of each phosphate group is reduced so that the total charge of the duplex is -5. Because temperature replica exchange MD (T-REMD) simulations on gas-phase duplexes had never been attempted before, LC and DC models were both tested for T-REMD.

Solution and gas phase MD simulations were carried out with the MPI-versions of modules *pmemd* and *sander*, respectively, of the Amber12 suite of programs,²⁰ implementing parmBSC1 force field²¹ for DNA and parmBSC0 force field + χ_{OL3} correction^{22,23} for RNA. The electrostatic interactions were calculated with the particle mesh Ewald algorithm²⁴ (real-space cut-off = 10 Å) in solution and direct Coulomb summation (no cut-off) in gas phase. All 12-bp duplexes were subjected gas phase T-REMD²⁵ (1 μ s \times 18 replicas) with temperature values from 300.00 to 633.94 K, chosen with predictor from reference 26 (average successful exchange rate of ca. 30%). Short duplexes (7—10 bp) were optimized at DFT level,²⁷ with the M06-2X²⁸ functional including the dispersion correction GD3.²⁹ The basis set was 6-31G(d,p) for the 7—9-bp duplexes (see Supporting Information) and 4-31G(d) for the 10-bp duplex. Duplexes up to 12-bp were also studied at the semi-empirical (SE) level with MOPAC,³⁰ using different methods³¹ (further details in Supporting Information). Hydrogen bond and stacking analysis was performed for all simulations as detailed in the Supplementary Information section S2.

Simulation of desolvation and proton transfer. Starting from equilibrated MD simulations in solution, we cut droplets of ca. 2400 water molecules (radius \sim 25 Å) containing the duplex 12-d₆₆, and 16, 17 and 18 NH₄⁺ cations to give a total net charge of -6, -5 and -4 respectively. The droplets were then subjected to gas-phase MD simulations following Konermann's protocol.³² Briefly, the trajectories were propagated by 500-ps chunks at constant temperature (350 K). To accelerate the evaporation, at the beginning of each chunk the initial velocities were reassigned according to the Boltzmann distribution at T = 350 K. At the end of each chunk, we stripped out all water molecules farther than 60 Å from the N6 atom of the 18th residue adenine (which is approximately in the center of the duplex). A further 50-ns chunk at T = 450 K helped the last “sticky” water molecules to evaporate. In total, twelve

independent trajectories were obtained (four at each charge state, 4-, 5- and 6-). We then localized the charges (LC model) as follows: on the ultimate conformation of every trajectory a proton from each NH₄⁺ cation was transferred to the closest phosphate oxygen atom, and ammonia is removed. The resulting duplexes were then subjected to 1- μ s unbiased MD.

CCS calculations. The collision cross section (CCS) is calculated using the EHSSrot code³³ with the atom parameterization of Siu et al.,³⁴ a combination that is both accurate and efficient for calculating the CCS of nucleic acids in the gas phase.³⁵ The CCS is calculated for snapshots every 0.5 ns in each MD trajectory.

RESULTS AND DISCUSSION

1) DNA duplexes in their predominant native charge state are more compact than a canonical B-helix.

When sprayed from aqueous 100 mM NH₄OAc, the major charge state of 12-bp duplexes is 5- (Figure S2). The ^{DT}CCS_{He} distributions for the 5- duplexes without ammonium ions bound and in the softest ion transfer conditions are shown in Figure 1A (full results in supporting Figure S3). These distributions suggest more compact structures than those of canonical B- or A-forms (Table 2). Upon pre-IMS activation, the CCS distribution of 12-d₃₃ is unchanged, 12-d₆₆ is losing the high-CCS peak to the profit of the low-CCS peak, and the entire distribution of 12-d₁₀₀ is shifting towards lower CCS, down to \sim 705 Å² (see supporting information Figure S4). At low pre-IMS activation, duplexes⁵⁻ with ammonium adducts are also detected. The CCS of these ions is similar to that of the bare duplexes, but for 12-d₆₆ the higher-CCS peak is more abundant—see supporting Figure S5). The peak center values obtained for soft and harsh conditions are listed in Table 1.

Charge states 4- and 6- are also detected in aqueous NH₄OAc. The duplex CCS distributions are highly charge-state dependent (see supplementary Figure S3): charge states 4- and 5- are similarly compact, whereas charge state 6- has a \sim 20% larger CCS. Charge state 7-, obtained for 12-d₆₆ and 12-d₁₀₀ by adding the “super-charging” agent sulfolane,³⁶ has a $>$ 30% larger CCS.

The duplex CCS distributions are significantly broader than those of the tetramolecular G-quadruplex [dTG₄T]₄ (in black in Figure 1A, and supplementary Figure S3), a rigid structure of the same size. This indicates that a greater conformational space is explored in the gas phase by nucleic acid duplexes compared to the G-quadruplex,¹⁴ and that gas-phase duplexes consist of an ensemble of conformations not fully interconverting on the time scale of the mobility separation (10–30 ms).

2) MD trajectories, DFT and semi-empirical optimizations reveal phosphate-phosphate hydrogen bond formation.

To find out which tridimensional structures are compatible with the experimental CCS values of the 12-bp⁵⁻ duplexes, we first carried out unbiased MD simulations directly from B- and A-helix structures, stripped of the solvent. Figure 1B,C shows representative CCS distributions. All such simulations give CCS values significantly larger than the experimental ones. Independent trajectories, started from different conformations and with different protonation states (*i.e.*, different sets of localized charges), confirmed this result (supporting Figure S7). The experimental CCS values of the duplexes⁶⁻ are nonetheless compatible with the simulated A- and B-helices, and duplexes⁷⁻ are compatible with B-helices. This does not necessarily mean that the gas-phase conformations of 6- and 7- charge states *are* A- and B-helices, but helps to understand conclusions previously derived solely based on more densely charged duplexes.^{17,18}

When starting from the B-form, MD simulations always show spontaneous hydrogen bonds formation between phosphate groups situated on each side of the minor groove (supporting information Movie S1 and Figure S8). This causes the “zipping” of the minor groove. The structures are stable up to 1 μ s (Figure S8), but too elongated compared to the duplexes⁵⁻ experimental data (compare Fig. 1B with 1A). In simulations starting from the A-helix, zipping occurs as well, but this time across the major groove (supporting Figure S9). In the gas phase, the closest protonated phosphate groups therefore tend to form hydrogen bonds that did not exist in solution. However, this kind of trajectory does not reflect an experimentally relevant one.

To check for possible artifacts due to using classical force fields to represent macromolecules in the gas phase, we also studied B-duplexes at the density functional theory (DFT) and semi-empirical (SE) levels. Upon DFT optimization of 7-bp to 10-bp duplexes, phosphate groups form new H-bonds and close the minor groove as well (Figure 2 and Supporting Figure S11). *WC* H-bonds and base stacking are well preserved,^{37,38} and the helix compresses along its longitudinal axis (Figure 2C) and the ^{CALC}CCS_{He} of the DFT optimized structure is smaller than that of the canonical helix.

For SE calculations, we first validated that the PM7 method best reproduces the DFT results (supporting Figure S12). The CCS values obtained after PM7 optimization are summarized in Table 2 for all [12-bp duplexes]⁵⁻ (A- and B-form for DNA and A-form for RNA). Upon PM7 optimization, the duplex [12-d₆₆]⁵⁻ undergoes minor groove zipping, while *WC* H-bonds and base pair stacking interactions are preserved (Figure S13). The compaction compared to the solution structure is only ~8%, thus still far from the experimental value.

Thus, starting from naked canonical structures, neither geometry optimization nor unbiased MD trajectories lead the system towards the experimentally observed conformational ensemble. So, either the sampling is incomplete (simulated and experimental time scales differ), or the starting structures, obtained by desolvating and charging the duplexes all at once, inadequately reflect the electrospray droplet desolvation and declustering.

3) Temperature replica exchange MD (T-REMD) exploration of gas-phase conformational landscapes do not converge to the experimental structure.

To enhance sampling, we carried out T-REMD simulations on the 12-bp duplexes⁵⁻ starting from their canonical structures. We used both DC and LC models. DNA and RNA results, including representative snapshots, total hydrogen bonding, *WC* hydrogen bonding and stacking³⁹ occupancies, are summarized in supporting information Section S5 (Figures S14—S24). For RNA, the theoretical CCS distributions, both with the DC and

LC models, match with the experimental ones (Figures S15 and S18). However, the hydrogen bonding and stacking patterns (Figures S19-S21) always become scrambled in the gas phase. For DNA, the T-REMD results depend more critically on the charge location model: $^{CALC}CCS_{He}$ values agree with the experiments for the DC model (Figure 1D) and for *some* LC models (Figure 3).

With the DC model, the CCS values are closer to the experimental ones, yet still significantly too large for 12-ds₁₀₀⁵⁻. The GC base pairs are mostly preserved (Figure S36), but the duplex 12-d₃₃⁵⁻ is mostly melted (separated into single strands; see Figure S14). DC models lack the explicit protons and therefore cannot form phosphate-phosphate hydrogen bonds. As a result, the strands can dissociate upon T-REMD. The rate of strand dissociation occurrence ranks 12-d₃₃ > 12-d₆₆ > 12-d₁₀₀ (supporting Figure S14), in line with the relative gas-phase kinetic stabilities in tandem mass spectrometry.¹¹⁻¹⁴

When LC models are used, the T-REMD final structures depend significantly on the choice of charge location (Figure S17, S19—S21). For example, for [12-d₁₀₀]⁵⁻, a first model (Figure 3B) gives CCS values matching well with the experiment, but has lost most *WC* H-bonds, whereas a second model (Figure 3C) preserves *WC* H-bonds but its CCS values are much larger than the experimental ones (the representative structure resembles those obtained with unbiased simulations starting from the B-form).

In summary, the T-REMD results do not account simultaneously for the preservation of GC base pairs and the experimental collision cross sections. Yet they teach us that the duplex⁵⁻ conformations are closer to a compact globular shape than to the helices obtained by unbiased MD trajectories, and suggest that sampling problems are at least partially responsible of the unfitting of simulation and experimental CCSs.

4) Progressive duplex desolvation leads to experimentally relevant structures.

All simulations presented to this point, and nearly all previous simulations are assuming an instantaneous transfer of the duplex from solution to the gas phase.

But in practice, desolvation and declustering proceeds gradually during the experiment. Consta and co-workers⁴⁰ modeled the desolvation process of duplex dA₁₁•dT₁₁ at atomistic level in a water droplet containing Na⁺ and Cl⁻, and found that the duplex collapses inside the droplet when the Na⁺ cations are numerous enough to interact with the phosphates and reduce the size of both grooves. Interestingly, the resulting charge density is similar to those observed experimentally from native conditions. It is therefore likely that the DNA duplexes are desolvated via the charged residue model^{41,42} and that the compaction results from the association of NH₄⁺ cations to both minor and major groove before full desolvation. As a result, the starting structure for gas phase simulations might be quite different to the canonical A- or B- helices. Because electrostatic interactions prevail, the gas phase conformational space is very stiff, and an incorrect starting structure can significantly bias the entire trajectory.

Here we simulated the desolvation of duplex 12-d₆₆ placed in a droplet containing ~2400 water molecules and 17 NH₄⁺ (net charge = -5). When the simulation reaches 39.5 ns, 87 water molecules stick on the duplex together with the 17 NH₄⁺ cations. The simulation was pursued for 50 ns at 450 K, to allow this ultra-stable inner solvation shell of water molecules to evaporate. The ammonium ions sticking to the duplex are mainly located close to phosphate groups, often in-between them. Several trajectories (see 4-, 5- and 6- charge states in supporting information Section S6; Figures S25—S30) lead to a chain of ammonium ions in the minor groove. For the 5- charge state, ammonium ions are more distributed across both grooves, and thus upon evaporation both grooves get narrower to enable phosphate-ammonium-phosphate salt bridges to form. Accordingly, the CCS value diminishes (Figure 4).

Classical MD cannot model proton transfer, but if water evaporation is almost complete before the proton transfers start, then the structures generated by gradual desolvation are good candidate for modeling the electrosprayed structures. Also, ammonium ion positioning upon desolvation could predict which phosphate groups will share a proton and form hydrogen bonds after complete declustering.

To simulate the eventual ammonia loss, we arbitrarily transferred a proton from each ammonium ion to its closest phosphate oxygen. The resulting desolvated and declustered duplex (12-d₆₆)⁵⁻ is stable over 1-μs MD at 300 K (Figure 4). The total hydrogen bond occupancy reveals additional contacts across both grooves. Remarkably, the CCS value now matches the experimental one, and at the same time, the generated structure keeps partial memory of the *WC* base pairs (at least, preserving mostly the GC ones), in line with CID and IRMPD results.

In summary, progressive desolvation, allowing the duplex to form phosphate-phosphate hydrogen bonds across both grooves can account simultaneously for the experimental compactness and for partial preservation of GC base pairs. In contrast, upon unbiased MD or structure optimization (DFT or SE), only the phosphate groups that were the closest in the starting structure (across the minor and major groove for B- and A-helices, respectively) could mate.

5) Longer duplexes, DNA and RNA alike, undergo compaction when electrosprayed from aqueous solutions of physiological ionic strength.

Past studies on GC-rich DNA duplexes, with one charge every two base pairs, had suggested that A-helices predominate from 8-bp on, and that B-helices are preserved from 18-bp on. To ascertain whether the compaction of lower charge states is a general phenomenon, we measured 12-bp to 36-bp DNA and RNA duplexes, from native solutions, either 100% aqueous or containing sulfolane. The tested duplexes included multiples of d₆₆, r₆₆, d₁₀₀ and r₁₀₀, and the DNA duplexes listed as potential CCS calibrants by the Fabris group.⁴³ The results are shown in Figure 5 (experimental values in supporting Table S3).

We found that DNA and RNA duplexes have very similar gas-phase CCS values, although their helix types differ in solution (B vs. A). The theoretical CCS values obtained with unbiased *in-vacuo* MD directly from the solution B-helix are overlaid (B-helix MD trend line in Fig. 5). They match only with some of the high charge states produced in the presence of sulfolane. However, the low charge states produced from purely aqueous NH₄OAc solutions are significantly more compact.

To reproduce the dual-groove closing of longer duplexes while avoiding computationally costly desolvation simulations, we opted for biased exploration. We used restrained minimization and imposed distance constraints based on our knowledge that compact structures can form via hydrogen-bond formation between phosphates across both grooves. Distance restraints were imposed using a harmonic potential between the hydrogen atom of neutral phosphates belonging to one strand and the oxygen atom of mating phosphates belonging to the other strand (Supporting Information Section S8, Figures S31—S33, and Movie S2 for a 12-bp duplex). The systems were minimized *in vacuo*, then all restraints were removed prior to 1-μs gas phase MD.

For 12-bp duplexes, the resulting CCS distributions agree with the experiment, with the strongest compaction for 12-d₁₀₀ (Figure 1E). Applying the same procedure starting from an A-helix however did not lead to similarly low CCS values (supporting Figure S34). The doubly groove-zipped helices obtained by restrained minimization generally keep *WC* hydrogen bonds less well than those obtained by desolvation (supporting Figure S35—S37), but still reflect the solution trend, with the 12-d₁₀₀ preserving the highest fraction of hydrogen bonds. The advantage of the procedure is to reproduce the phosphate-phosphate H-bond pattern of the desolvated helices (two diagonals for zipping across both grooves, plus the central diagonal indicating preserved base pairs, Figure S35). We then applied restrained minimization to the longer (24-bp and 36-bp) helices (supporting Figures S38—S39). Whatever the duplex length, the experimental CCS values obtained for low charge states (Figure 5) match better with these zipped helices than with the canonical structures or with the helices relaxed by long unbiased gas-phase MD.

CONCLUSIONS

In summary, at the charge states produced by electrospray from aqueous ammonium acetate (traditional “native” solution conditions), double-stranded nucleic acids undergo a significant compaction in gas phase compared to the structure in solution. Unbiased molecular dynamics of B-helix or A-helix structures directly

transposed from solution to the gas phase fails to reproduce the experimental results. This is due to several reasons: i) only the phosphate groups closest to one another can pair to form hydrogen bonds on the simulation time scale, ii) the starting structure is unrealistic and iii) sampling in unbiased MD simulations is intrinsically limited. In the case of T-REMD, depending on the initial choice of charge location, the final structures either did not have any memory of the solution structure, or resembled those obtained by unbiased MD. T-REMD can help to solve the sampling effect (the question of the maximum internal temperature reached in the experiments remaining open), but not the problem that original charge locations might be incorrect.

Gradual desolvation generates more realistic starting structures for gas phase simulations. Conformational transitions occurring during dehydration cannot be ignored because they guide the entire sampling, within a particularly stiff conformational landscape in the case of nucleic acids. The conformationally restrained duplexes remain stable upon unbiased MD: once formed, they stay locked at room temperature. The broadness of the experimental CCS distributions therefore indicates a distribution of co-existing—but not interconverting—conformations, wherein each would have a slightly different phosphate-phosphate hydrogen bond network.

Our results highlight a key difference between nucleic acids and proteins native mass spectrometry. Globular proteins can rearrange by relaxing their side chains⁴⁴ and undergo minimal salt bridge rearrangement.⁴⁵ Briefly optimized structures often have CCS values matching well with the experiments.^{42,46-48} Fabris and co-workers have recently underlined the difficulties in transposing to DNA the MD and CCS calculation protocols traditionally used for proteins.⁴³ As a way out they proposed to calibrate all traveling wave IMS data using short MD simulation results, but our study shows why this approach would lead to a misrepresentation of nucleic acid structures in the gas phase. DNA and RNA double helices are more compact in the gas phase than in solution, due mostly to new phosphate-phosphate interactions. At the low charge states produced from ammonium acetate, the Coulomb repulsion is not sufficient to keep the phosphate groups apart. They rearrange by self-solvation, cause major rearrangements of

the backbone, and lead to a significant compaction (>20%) compared to the starting structure. Yet, they are metastable conformations keeping some memory of what the structure was just before vaporization.

SUPPORTING INFORMATION

Experimental procedures including reconstruction of the CCS distributions, full ESI-IMS-MS results, detailed computational procedures, full modeling results by *in vacuo* QM optimization, MD, T-REMD, and MD following restrained minimization, structural analysis, and supplementary results on longer duplexes (PDF).

Movie S1: minor groove zipping upon MD of a 12-bp B-helix (GIF)

Movie S2: dual groove zipping imposed by restrained minimization on a 12-bp B-helix (GIF).

ACKNOWLEDGMENTS

This work was funded by the European Research Council (ERC DNAFOLDIMS to VG and ERC SimDNA to MO), by the Spanish BIO2015-69802-R grant to MO, and by EU COST action BM1403 (short-term scientific mission to MP). HG is a Juan de la Cierva researcher.

We are thankful to David E. Condon for donating us the Perl script used to score the base stacking, and to Adrien Marchand and members of COST action BM1403 for fruitful discussions. LD is a SNI (Sistema Nacional de Investigadores; ANII, Uruguay) researcher, and MO is an ICREA-academia fellow.

ABBREVIATIONS

bp, base pair; CCS, collision cross section; ^{CALC}CCS_{He}, calculated CCS in helium; ^{DT}CCS_{He}, CCS in helium measured in a drift tube; DC, distributed charges; DFT, density functional theory; ESI, electrospray ionization; GC, guanine-cytosine; IMS, ion mobility spectrometry; LC, localized charges; MD, molecular dynamics; MS, mass spectrometry; SE, semi-empirical; T-REMD, temperature replica exchange molecular dynamics; *WC*, Watson-Crick.

REFERENCES

- (1) Blackburn, G. M.; Gait, M. J. *Nucleic acids in chemistry and biology*; Oxford University Press: Oxford, 1996.
- (2) Rhodes, D.; Lipps, H. J. *Nucleic Acids Res.* **2015**, *43*, 8627.
- (3) Breaker, R. R. *Mol. Cell* **2011**, *43*, 867.
- (4) Loo, J. A. *Mass Spectrom. Rev.* **1997**, *16*, 1.
- (5) Breuker, K. *Int. J. Mass Spectrom.* **2004**, *239*, 33.
- (6) Rosu, F.; De Pauw, E.; Gabelica, V. *Biochimie* **2008**, *90*, 1074.
- (7) Kitova, E. N.; El-Hawiet, A.; Schnier, P. D.; Klassen, J. S. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 431.
- (8) Zheng, X.; Liu, D.; Klarner, F. G.; Schrader, T.; Bitan, G.; Bowers, M. T. *J. Phys. Chem. B* **2015**, *119*, 4831.
- (9) Canon, F.; Ballivian, R.; Chiro, F.; Antoine, R.; Sarni-Manchado, P.; Lemoine, J.; Dugourd, P. *J. Am. Chem. Soc.* **2011**, *133*, 7847.
- (10) Hyung, S. J.; Robinson, C. V.; Ruotolo, B. T. *Chem. Biol.* **2009**, *16*, 382.
- (11) Schnier, P. D.; Klassen, J. S.; Strittmatter, E. F.; Williams, E. R. *J. Am. Chem. Soc.* **1998**, *120*, 9605.
- (12) Gabelica, V.; De Pauw, E. *Int. J. Mass Spectrom.* **2002**, *219*, 151.
- (13) Pan, S.; Sun, X. J.; Lee, J. K. *Int. J. Mass Spectrom.* **2006**, *253*, 238.
- (14) Burmistrova, A.; Gabelica, V.; Duwez, A. S.; De Pauw, E. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 1777.
- (15) Gabelica, V.; Rosu, F. In *Nucleic Acids in the Gas Phase*; Gabelica, V., Ed.; Springer: Berlin, 2014, p 103.
- (16) Rueda, M.; Kalko, S. G.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2003**, *125*, 8007.
- (17) Gidden, J.; Ferzoco, A.; Baker, E. S.; Bowers, M. T. *J. Am. Chem. Soc.* **2004**, *126*, 15132.
- (18) Baker, E. S.; Bowers, M. T. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1188.
- (19) Macke, T.; Case, D. A. In *Molecular Modeling of Nucleic Acids*; Leontes, N. B., SantaLucia, J., Jr., Eds.; American Chemical Society: Washington, DC, 1998, p 379.
- (20) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.

Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A.; University of California, San Francisco: 2012.

(21) Ivani, I.; Dans, P. D.; Noy, A.; Perez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goni, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpi, J. L.; Gonzalez, C.; Vendruscolo, M.; Laughton, C. A.; Harris, S. A.; Case, D. A.; Orozco, M. *Nat. Methods* **2016**, *13*, 55.

(22) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817.

(23) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E., 3rd; Jurecka, P. *J. Chem. Theory Comput.* **2011**, *7*, 2886.

(24) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

(25) Sugita, Y.; Okamoto, Y. *Chem. Phys. Chem.* **1999**, *314*, 141.

(26) Patriksson, A.; van der Spoel, D. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2073.

(27) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; Gaussian, Inc.: Wallingford, CT, USA, 2009.

(28) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *120*, 215.

(29) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.

(30) Stewart, J. J. P.; Stewart Computational Chemistry: Colorado Springs, CO, USA, 2012.

(31) Stewart, J. J. *J. Mol. Model.* **2013**, *19*, 1.

(32) Metwally, H.; McAllister, R. G.; Konermann, L. *Anal. Chem.* **2015**, *87*, 2434.

(33) Shvartsburg, A. A.; Mashkevich, S. V.; Baker, E. S.; Smith, R. D. *J. Phys. Chem. A* **2007**, *111*, 2002.

(34) Siu, C. K.; Guo, Y.; Saminathan, I. S.; Hopkinson, A. C.; Siu, K. W. *J. Phys. Chem. B* **2010**, *114*, 1204.

(35) D'Atri, V.; Porrini, M.; Rosu, F.; Gabelica, V. *J. Mass. Spectrom.* **2015**, *50*, 711.

(36) Ogorzalek Loo, R. R.; Lakshmanan, R.; Loo, J. A. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 1675.

(37) Zubatiuk, T.; Kukuev, M. A.; Korolyova, A. S.; Gorb, L.; Nyporko, A.; Hovorun, D.; Leszczynski, J. *J. Phys. Chem. B* **2015**, *119*, 12741.

- (38) Zubatiuk, T. A.; Shishkin, O. V.; Gorb, L.; Hovorun, D. M.; Leszczynski, J. *Phys. Chem. Chem. Phys.* **2013**, *15*, 18155.
- (39) Condon, D. E.; Kennedy, S. D.; Mort, B. C.; Kierzek, R.; Yildirim, I.; Turner, D. H. *J. Chem. Theor. Comput.* **2015**, *11*, 2729.
- (40) Sharawy, M.; Consta, S. *Phys. Chem. Chem. Phys.* **2015**, *17*, 25550.
- (41) Dole, M.; Mack, L. L.; Hines, R. L. *J. Chem. Phys.* **1968**, *49*, 2240.
- (42) McAllister, R. G.; Metwally, H.; Sun, Y.; Konermann, L. *J. Am. Chem. Soc.* **2015**, *137*, 12667.
- (43) Lippens, J. L.; Ranganathan, S. V.; D'Esposito, R. J.; Fabris, D. *Analyst* **2016**, *141*, 4084.
- (44) Breuker, K.; McLafferty, F. W. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 18145.
- (45) Loo, R. R.; Loo, J. A. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 975.
- (46) Pacholarz, K. J.; Porrini, M.; Garlish, R. A.; Burnley, R. J.; Taylor, R. J.; Henry, A. J.; Barran, P. E. *Angew. Chem. Int. Ed.* **2014**, *53*, 7765.
- (47) Meyer, T.; de la Cruz, X.; Orozco, M. *Structure* **2009**, *17*, 88.
- (48) Harvey, S. R.; Porrini, M.; Konijnenberg, A.; Clarke, D. J.; Tyler, R. C.; Langridge-Smith, P. R.; MacPhee, C. E.; Volkman, B. F.; Barran, P. E. *J. Phys. Chem. B* **2014**, *118*, 12348.
- (49) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33.

Table 1. Size, name and sequences of the duplexes under study, and outline of calculations carried out.

Size	Name	Sequence	Theoretical levels of study
10 bp	10-bp	dCGCGGGCCCCG•dCGGGCCCCGCG	DFT (M06-2X - 4-31G(d)) semi-empirical (PM7)
12 bp	12-d ₃₃	(dCGTAAATTTACG) ₂	semi-empirical (PM7)
	12-d ₆₆	(dCGCGAATTCGCG) ₂	MD from different starting structures (1 μs each)
	12-d ₁₀₀	(dCGCGGGCCCCGCG) ₂	T-REMD (1 μs × 18 replicas)
	12-r ₃₃	(rCGUAAAUUUACG) ₂	semi-empirical (PM7) T-REMD (1 μs × 18 replicas)
	12-r ₆₆	(rCGCGAAUUCGCG) ₂	
	12-r ₁₀₀	(rCGCGGGCCCCGCG) ₂	
24 bp and 36 bp	Concatenations of the 12-bp duplexes above. See supporting information Section S7.		MD on B-helix (0.25 μs) MD on zipped helix (0.5 μs)

Table 2. Collision cross section values in helium (CCS_{He}) of 12-bp duplexes, prior and subsequent to PM7 optimization, unbiased MD, T-REMD (DC model; results with LC model are not listed because they depend on the chosen locations—see text), and unbiased MD on the helix zipped by restrained minimization. Errors reported for MD are the standard deviation on the different structures along the trajectory.

	$\text{DTCCS}_{\text{He}} (\text{\AA}^2)$	$\text{CALCCCS}_{\text{He}} (\text{\AA}^2)$								
		B-helix				A-helix				Zipped helix
		Canonical	PM7	MD	T-REMD (DC)	Canonical	PM7	MD	T-REMD (DC)	MD
$[12\text{-d}_{33}]^{5-}$	Soft/harsh: 760	918	914	954 ± 18	780 ± 24	900	861	826 ± 11		759 ± 12
$[12\text{-d}_{66}]^{5-}$	Soft: 735/803 Harsh : 735	903	831	881 ± 6	770 ± 15	893	942	886 ± 10		752 ± 10
$[12\text{-d}_{100}]^{5-}$	Soft: 730 Harsh: 710	908	838	953 ± 8	771 ± 14	892	869	774 ± 6		711 ± 9
$[12\text{-r}_{33}]^{5-}$	-	-	-			944	938		741 ± 21	
$[12\text{-r}_{66}]^{5-}$	Soft/harsh: 725	-	-			945	902		747 ± 18	
$[12\text{-r}_{100}]^{5-}$	Soft: 735 Harsh: 695	-	-			940	896		743 ± 11	

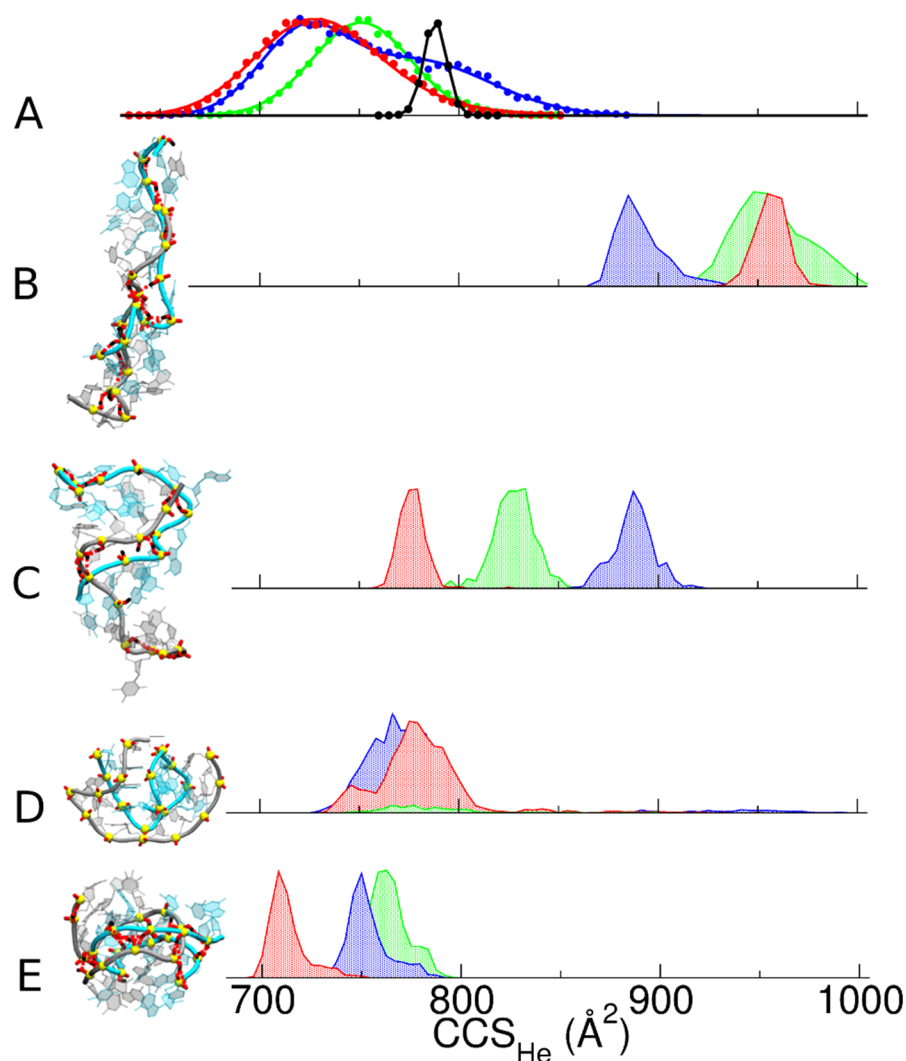


Figure 1. (A) Experimental $^{DT}CCS_{He}$ distribution for the 12-bp DNA duplexes (green: (12-d₃₃)⁵⁻, blue: (12-d₆₆)⁵⁻, red: (12-d₁₀₀)⁵⁻) and the rigid G-quadruplex [d(TG₄T)]₄⁵⁻. (B-E) Calculated CCS distributions for molecular models generated by (B) gas-phase MD of the B-helices, (C) gas-phase MD of the A-helices, (D) T-REMD simulations on B-helices using distributed charges (note that most of the population of (12-d₃₃)⁵⁻ duplex dissociated during simulations), and (E) MD following a restrained minimization forcing H-bond formation between the phosphate groups across both grooves of a B-helix. The final MD structures of each duplex model, created with VMD software,⁴⁹ are shown for (12-d₆₆)⁵⁻ on the same scale (see supporting Figure S10 for a magnification).

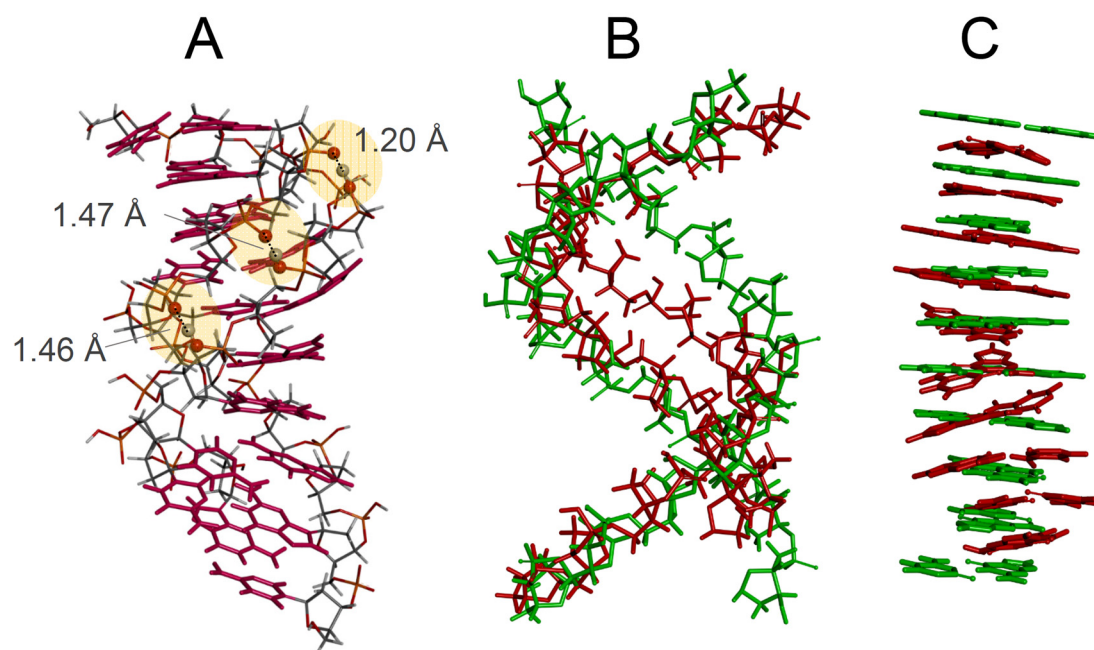


Figure 2. DFT optimization of the 10-bp duplex⁴⁻: (A) new hydrogen bonds formed between phosphate groups across the minor groove; (B) superposition of the sugar-phosphate backbones of the canonical B-helix (green) and optimized duplex (red); (C) superposition of the base pairs of the canonical B-helix (green) and optimized duplex (red).

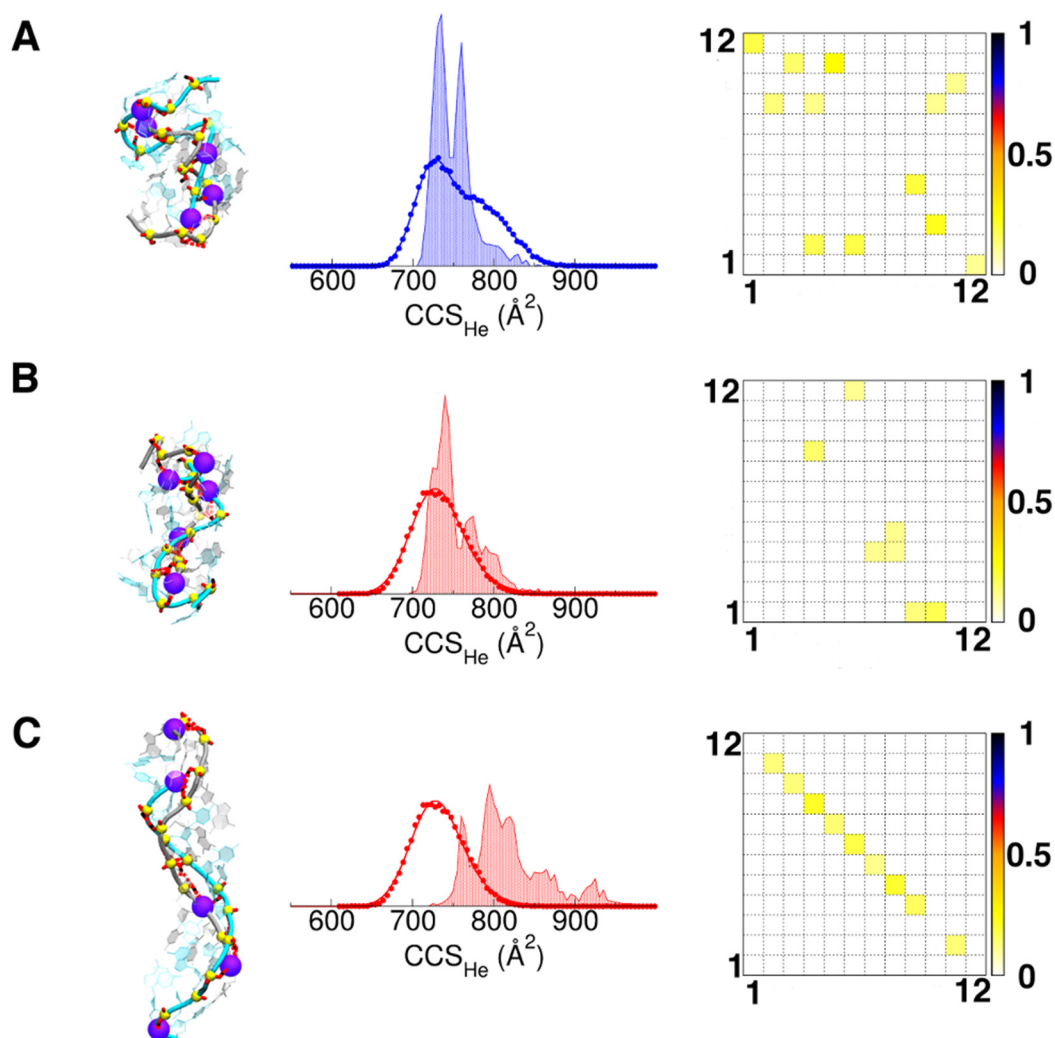


Figure 3. T-REMD on 12-bp DNA duplexes⁵⁻ with localized charges (LC). (A) [12-d₆₆]⁵⁻, (B-C) [12-d₁₀₀]⁵⁻ with different LC models. The location of negatively charged phosphates is shown by the beads on the representative snapshot structure. The experimental ^{DT}CCS_{He} distribution and calculated one (shade) are overlaid. The fractional WC H-bond occupancies are shown on the right: the bases of each strand are numbered from 5' to 3', and perfect base pair matching is therefore indicated by a diagonal from top left to bottom right.

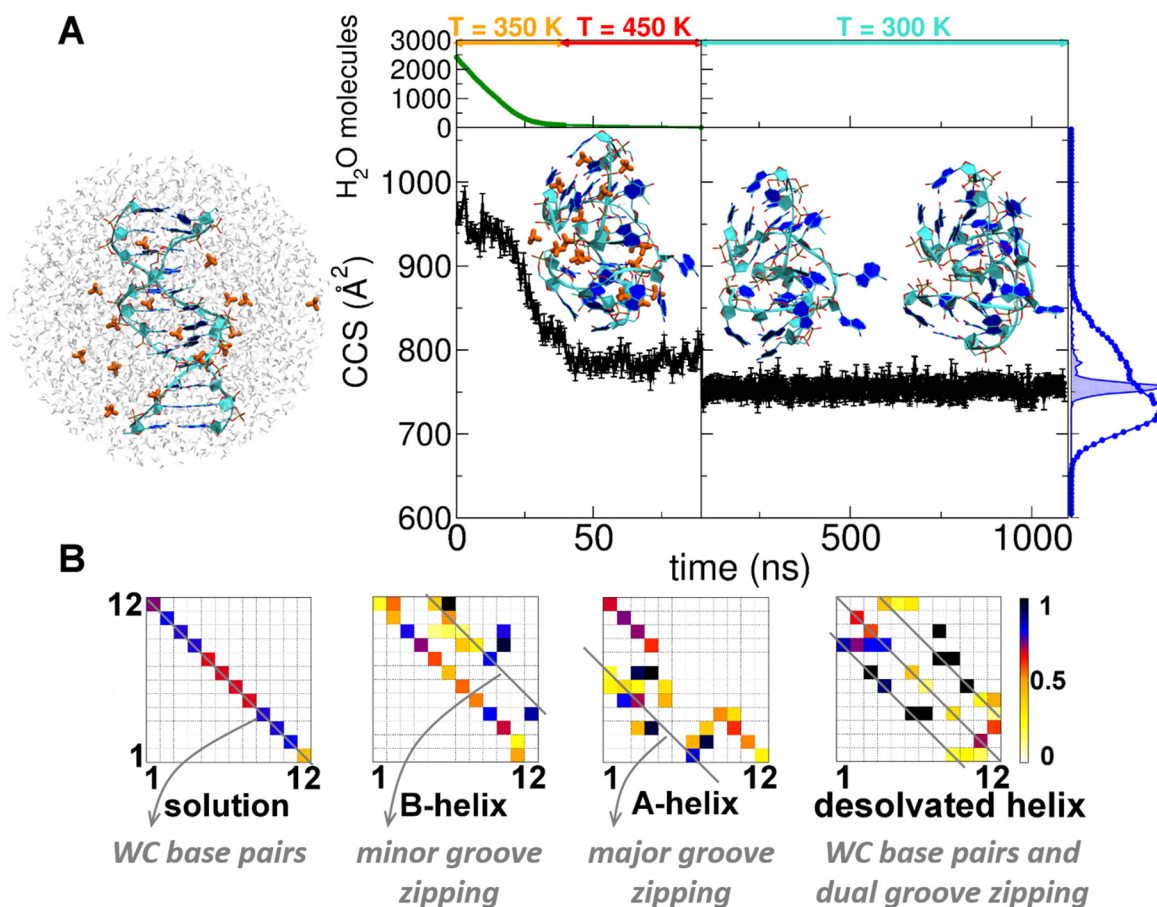


Figure 4. (A) $^{CALC}CCS_{He}$ evolution during the desolvation of the 12-d₆₆ droplet with -5 net charge. At the top the simulation temperatures are shown along the related trajectory portions. The initial, intermediate and final structures are shown with DNA strands in cyan, bases aromatic rings in blue and NH₄⁺ cations in orange. On the right the [12-d₆₆]⁵⁻ $^{DT}CCS_{He}$ distribution (blue circles) is superimposed on the $^{CALC}CCS_{He}$ (blue area) one. (B) H-bond occupancies of the desolvated helix (right), compared to Watson-Crick H-bond occupancies of a B-helix in solution (left), and to gas-phase B- or A- helices upon unbiased MD in the gas phase (middle). In the B-helix, extra H-bonds form between phosphates across the minor groove; in the A-helix, across the major groove, and in the desolvated helix, across both grooves.

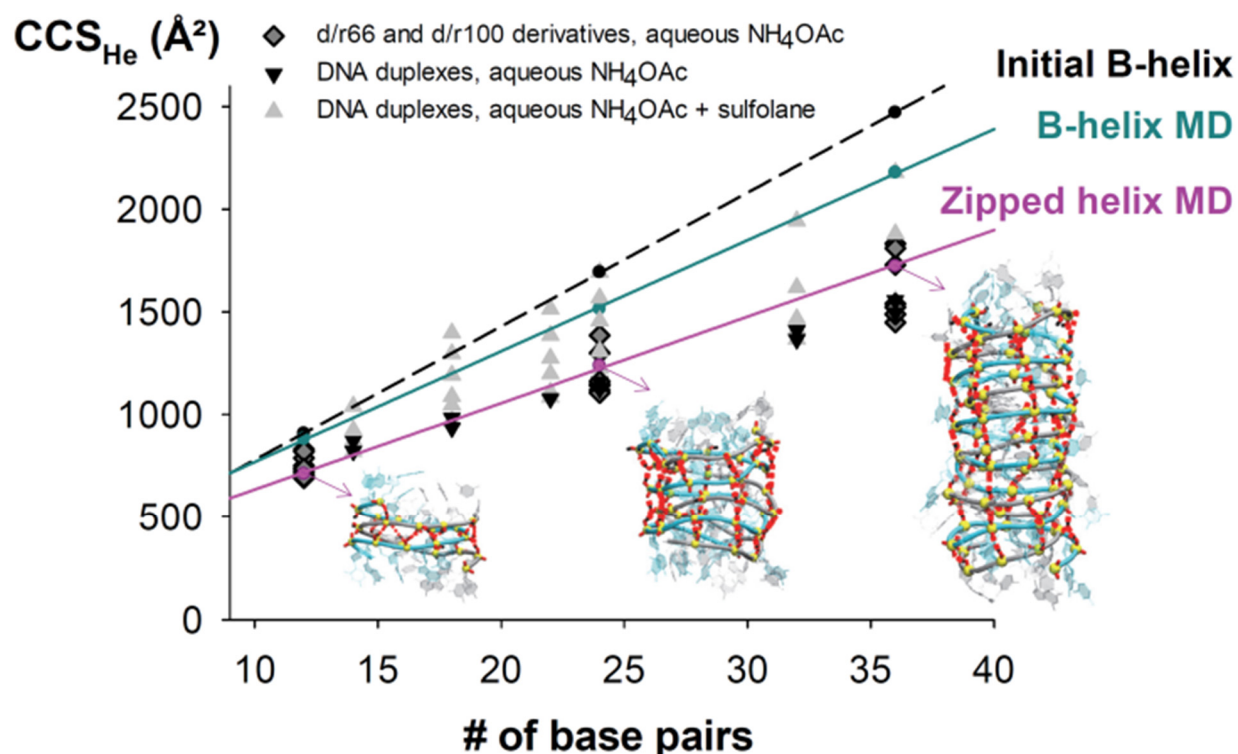


Figure 5. Comparison between experimental and calculated CCS_{He} of 12-bp to 36-bp duplexes. Diamonds and black down triangles: experiments on DNA and RNA duplexes in aqueous 100-mM NH₄OAc, respectively. These experimental points match better with MD on the zipped helix (restrained minimization followed by MD). DNA duplexes at higher charge states produced by sulfolane addition adopt more extended conformations, which match better with MD simulations on B-helices.